

## ALGORITMOS GENÉTICOS APLICADOS A LA RECUPERACIÓN DE INFORMACIÓN

Ana Isabel Pestana Caldes

*Escola Secundária/3 da Rainha Santa Isabel de Estremoz (Portugal)*

Cristina López-Pujalte

*Universidade de Extremadura (Espanha)*

### Introducción

El presente trabajo consiste en el análisis de investigaciones científicas acerca de técnicas de computación evolutivas [1, 3] aplicadas al ámbito de la Recuperación de Información (R.I.) [2, 4] durante el periodo de los últimos 20 años (1988-2008). Se analiza la producción científica existente en este campo, su evolución en el tiempo, principales investigadores, y sobre todo, las herramientas genéticas más utilizadas, las áreas específicas de aplicación dentro del ámbito de la R.I., así como los principales resultados obtenidos.

### Objetivos y Metodología

El principal objetivo del presente trabajo es obtener una visión general de las investigaciones realizadas con técnicas genéticas aplicadas al campo de la R.I. Para ello, el estudio se realiza sobre los documentos indizados en dos prestigiosas bases de datos: *Scopus* (<http://www.elsevier.com>) e *ISI WOS* (<http://www.thomsonreuters.com>). Se han seleccionado únicamente los documentos primarios (artículos y revisiones) en el periodo 1988-2008, ya que es a partir de ese año cuando aparecen las primeras aplicaciones de algoritmos genéticos al campo de la R.I., lo que nos proporciona un conjunto de 140 documentos sobre nuestro tema.

En primer lugar se ha realizado una búsqueda exhaustiva sobre ambas bases de datos, basada en palabras claves, con el objetivo de recuperar el mayor número posible de documentos, descartándose los que no pertenecían al tema objeto de estudio. Posteriormente, se ha recogido toda la información referente a cada documento en una base de datos. Cada documento tuvo que ser analizado en profundidad para determinar su área de aplicación, la herramienta utilizada, y los resultados alcanzados.

### Resultados y Discusiones

Si analizamos la evolución de la producción en el tiempo, es a partir del año 2003 cuando se produce un considerable aumento de las publicaciones, siendo el 2006, el año con más publicaciones científicas (Fig. 1).

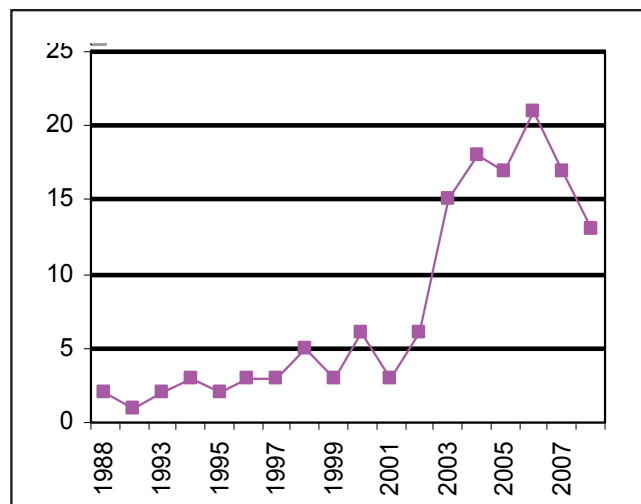


Fig. 1 – Número de publicaciones por año

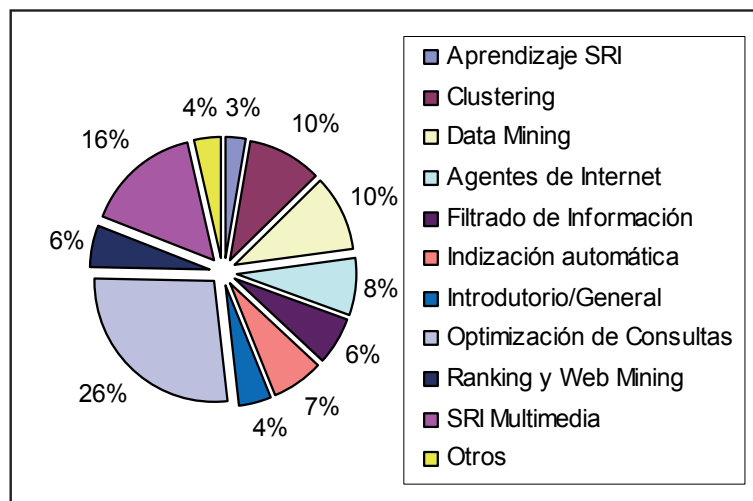


Fig. 2 – Clasificación de la producción

Los resultados obtenidos confirman los de un estudio previo [3] y muestran cuatro líneas fundamentales de aplicación dentro de este tipo de investigaciones: Optimización automática de consultas, Sistemas de Recuperación de Información Multimedia, Clasificación (clustering), y Data mining (éstas dos últimas a veces relacionadas), aunque también existen otras áreas de aplicación interesantes pero menos numerosas (Fig. 2).

Asimismo, se ha podido observar en el análisis, que las herramientas evolutivas más utilizadas son los Algoritmos Genéticos y la Programación Genética, y que la mayoría de las investigaciones muestran resultados muy prometedores en cada uno de los campos de aplicación.

En la Tabla 1 se recogen los principales autores (hemos considerado aquellos que tienen publicados más de 2 trabajos sobre el tema) y sus citas recibidas, destacando ampliamente *Chen*, y *Gordon*.

Con relación a las principales revistas (hemos considerado aquellas que tienen al menos 3 publicaciones), como se puede observar en la Tabla 2, destaca ampliamente “*Information Processing & Management*” sobre las demás, y son revistas con Factor de Impacto alto, siendo en su mayoría de las áreas *Computer Science: Information Systems*, y *Computer Science: Artificial Intelligence*.

Tabla 1 – Publicaciones y citas por autor

Autores	Publicações	Citações
Gordon, M. D.	8	174
Chen, H.	7	212
Fan, W.	7	89
Pathak, P.	7	89
Hirota, K.	6	48
Stejić, Z.	6	48
Takama, Y.	6	48
Moya-Anegón, F.	4	40
Guerrero-Bote, V.P.	3	40
López-Pujalte, C.	3	40
Boughanem, M.	3	28
Chrisment, C. Y.	3	28
Tamine, L.	3	28
Cordón, O.	3	10
Herrera-Viedma, E.	3	10

Tabla 2 – Total de publicaciones por revista y sus respectivas áreas

Revista	Publicações	Áreas
INFORM PROCESS MANAG	13	Computer Science, Information Systems
J AM SOC INF SCI	5	Computer Science, Information Systems
LECT NOTES COMP SCI	5	Mathematics: Theoretical Computer Science
EXPERT SYST APPL	4	Computer Science, Artificial Intelligence Engineering, Electrical & Electronic Operations Research & Management Science
INFORM SCIENCES	4	Computer Science, Information Systems Social Sciences Library and Information Science
DECIS SUPPORT SYST	3	Computer Science, Artificial Intelligence Computer Science, Information Systems Operations Research & Management Science
IEEE T EVOLUT COMPUT	3	Computer Science, Artificial Intelligence Computer Science, Theory & Methods
IEEE T SYST MAN CY B	3	Automation & Control Systems Computer Science, Artificial Intelligence Computer Science, Cybernetics
INFORM RETRIEVAL	3	Computer Science, Information Systems
J COMPUT INF SYST	3	Computer Science, Information Systems Engineering
J AM SOC INF SCI TEC	3	Computer Science, Information Systems

También se ha realizado una comparación de la producción científica contenida en las dos bases de datos del estudio, y aunque SCOPUS lógicamente presenta más documentos sobre el tema, ya que indiza mayor número de revistas que ISI WoS, ésta última tiene mayor precisión, presentando mucho menos ruido.

## Agradecimientos

Este trabajo ha sido financiado por la Junta de Extremadura-Vicepresidencia Segunda y Consejería de Economía, Comercio e Innovación y el Fondo Social Europeo, como parte del proyecto de investigación PRI06A233.

## Referencias bibliográficas

- [1] Bäck, T., Fogel, D.B., Michalewicz, Z. (1997). *Handbook of Evolutionary Computation*. New York: IOP Publishing and Oxford University Press.
- [2] Baeza-Yates, R. & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. New York: Addison Wesley.
- [3] Cordón, O., Herrera-Viedma, E., López-Pujalte, C., Luque, M., Zarco, C. (2003). A Review on the Application of Evolutionary Computation to Information Retrieval. *International Journal of Approximate Reasoning*, 34(2-3), 241-264.
- [4] Salton, G. & McGill, M.H. (1983). *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.

**Palabras-Clave:** Genetic Algorithms, Information Retrieval, Evolutionary Computation, Information Science.